DOCUMENT RESUME

ED 274 671 TM 860 522

AUTHOR McLarty, Joyce R.; Rakow, Ernest
TITLE Low-Inference Observation in the Evaluation of Tennessee's Career Ladder Teachers.
PUB DATE Apr 86
NOTE 28p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 17-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Career Ladders; *Classroom Observation Techniques; Correlation; Effect Size; Elementary Secondary Education; *Evaluation Criteria; Evaluation Methods; Factor Structure; *Performance Factors; Scores; State Programs; *Teacher Evaluation
IDENTIFIERS Tennessee; *Tennessee Career Ladder Program

ABSTRACT

This paper explores the functioning of the low-inference observation system used as part of the Tennessee Career Ladder evaluation system, which provides teachers with the opportunity to qualify for additional pay on the basis of outstanding performance. The analysis of the functioning of this form is intended to provide the basis for the ongoing improvement of Career Ladder evaluations and to increase understanding of the role of low-inference observation in the evaluation of teachers. The data reported are taken from Career Ladder data from the 1984-85 evaluation, but the metric used is not the same as that used in the actual evaluations. A number of changes in instrumentation and scoring methods were made after the 1984-85 evaluations. Data are based on a random sample of 275 teachers statewide and on actual evaluations of 3,132 teachers in 1984-85. The following conclusions were made: (1) the choice of aggregation level determines the nature of the analytic outcome; (2) the choice of scoring method exerts a powerful influence; and (3) the potential of pattern scoring is matched by the difficulty of developing meaningful scoring formulae. An appendix includes the observation form. (JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED274671

Low-Inference Observation in The Evaluation of Tennessee's Career Ladder Teachers

Joyce R. McLarty

State of Tennessee

Department of Education

Ernest Rakow

Memphis State University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. R. McLarty

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

TM 860 522

Running Head: LOW-INFERENCE OBSERVATION

BEST COPY AVAILABLE

ABSTRACT

Low-Inference Observation in
the Evaluation of Tennessee's Career Ladder Teachers

These analyses were conducted as part of the ongoing study of the Tennessee Career Ladder evaluation system. This system provides teachers with the opportunity to qualify for additional pay on the basis of outstanding teaching performance. Teachers evaluated under this state-directed program are assigned to the higher levels of the Career Ladder based on the evaluation results and years of experience. They are also provided detailed diagnostic information and related career development opportunities regardless of the outcome of the evaluation.

Part of the evaluation on which the determination of teaching excellence is based is an in-class observation using a low-inference type of observation form. The analysis of the functioning of this form is intended to provide the basis for the ongoing improvement of Career Ladder evaluations, and to increase understanding of the role of low-inference observation in the evaluation of teachers.

Acknowledgments

The authors gratefully acknowledge the assistance and support of Dr. Carol Furtwengler, Assistant Commissioner for the Career Ladder Program; Dr. George Malo, Associate Assistant Commissioner for the Career Ladder Program; and Dr. Sara Strouss, Director of the State Testing and Evaluation Center.

Cautions

The data reported in these studies are taken from Career Ladder data from the 1984-85 teacher evaluation system. The intent of this paper is to explore the functioning of the low-inference observation system used in the evaluations, not to assess the performance of the Career Ladder Teachers. The metric used here is not the same as that used in the actual Career Ladder evaluations (for which a scale of 1-9 was used). In addition, a number of changes in instrumentation and scoring methodology were made after the 1984-85 evaluations, so generalizations from those data to the 1985-86 evaluations are inappropriate.

Low-Inference Observation in the Evaluation of Tennessee's Career Ladder Teachers

These analyses were conducted as part of the ongoing study of the Tennessee Career Ladder evaluation system. This system provides teachers with the opportunity to qualify for additional pay on the basis of outstanding teaching performance. Teachers evaluated under this state-directed program are assigned to the higher levels of the Career Ladder based on the evaluation results and years of experience. They are also provided detailed diagnostic information and related career development opportunities regardless of the outcome of the evaluation.

Part of the evaluation on which the determination of teaching excellence is based is an in-class observation using a low-inference type of observation form. The analysis of the functioning of this form is intended to provide the basis for the ongoing improvement of Career Ladder evaluations, and to increase understanding of the role of low-inference observation in the evaluation of teachers.

Perspective

The Tennessee Career Ladder observation form was developed after existing observation forms were evaluated and found to be unsuited to the evaluation of teachers under this objectives-based program. Because each objective to be evaluated under the Career Ladder program had been carefully defined and approved, an observation system keyed to the specific objectives was required. The system had to focus on the domains of instructional planning, teaching strategies, evaluation, and classroom management. Each domain was further defined by subcategories called indicators of effective performance.

In order to measure the domains and indicators, a number of instruments (observation, questionnaires, structured interview) were developed. The system was designed so that several different instruments would be used to assess each domain. Figure 1 shows which instruments are used to measure which domains (designated by Roman numerals) and indicators (designated by capital letters). Although the focus of this paper is on the observation form, it is important to remember that it is only one of the six instruments used in this evaluation system.

Figure 1

Data Sources for

COMPETENCIES AND INDICATORS

	Classroom Observation	Portfolio	Applicant Interview	Principal Interview	Peer Questionnaire	Student Questionnaire	Written Test
I. Prepares for instruction effectively.							
A. Establishes appropriate instructional goals and related objectives consistent with the curriculum.		X	X	X			
B. Prepares instructional plans and materials incorporating principles of effective instruction.	X	X	X	X			
C. Creates, selects or modifies instructional plans and materials to accommodate learner instructional levels.	X						
II. Uses teaching strategies and procedures appropriate to the content, objectives, and learners.							
A. Provides a clear description of the learning task and its content.	X	X				X	
B. Monitors learner understanding and reteaches as necessary.	X	X	X			X	
C. Provides learners appropriate practice and review.	X	X	X			X	
D. Establishes and maintains learner involvement in the learning task.	X		X			X	
III. Uses evaluation to improve instruction.							
A. Uses information about learner performance to improve the instructional process.		X	X	X			
B. Reports learner status and progress to learners and their parents.	X	X	X	X		X	
IV. Manages classroom activities effectively.							
A. Establishes and maintains appropriate learner behavior.	X	X		X		X	
B. Establishes and maintains a classroom climate conducive to learning.	X	X	X	X		X	
C. Makes effective use of classroom resources (e.g. personnel, time, materials, facilities)	X		X	X		X	
V. Establishes and maintains a professional leadership role.							
A. Improves professional skills and knowledge.		X	X	X			
B. Takes a leadership role in improving education.		X	X	X	X		
C. (Screening only) Performs professional responsibilities efficiently.					X		
VI. (Screening only) Communicates effectively.							
A. Writes clearly and correctly.	X					X	X
B. Communicates oral information effectively.	X					X	
C. Reads professionally relevant literature/materials with comprehension.							

BEST COPY AVAILABLE

Methods

The Career Ladder observation form was developed and field tested during 1983-84 and put into use in 1984-85 as part of the Career Ladder evaluation system. Each observation was approximately one hour long and was preceded by a pre-observation conference with the teacher to orient the evaluator to the classroom. Each observation was followed by a post-observation conference with the teacher to clarify any aspects of the class which concerned the evaluator or teacher and to provide feedback to the teacher.

The observation form itself consisted of four basic sections. A copy of these sections is included in the appendix. The first section (front cover) simply provided space for noting classroom context conditions and was generally completed before the start of the class. The second (even-numbered pages) was done in alternating four minute scans and provided for marking a variety of teacher behaviors (task and content presentation, monitoring, feedback, evaluation, etc.), classroom management information (students off task), classroom climate, and student interaction information. Section three (odd-numbered pages) was completed in alternating 2 minute scans and provides for recording additional information on classroom management and teaching strategies. Section four (back cover) was completed after each observation and provided a record of less transient aspects of teacher performance (use of facilities and equipment, communication skills, student preparedness).

Data was aggregated for each of the scans conducted during an observation (6 to 9 scans depending on the length of the observation). Scans were then combined by observation, and observations were combined by candidate. A total of 25 separate scores were generated for each observation, in addition to the more than 300 items in the raw data. Additional information collected for each candidate included the remaining items on the observation (see Appendix), background and demographic information on the teacher, and evaluation data from the rest of the Career Ladder instruments (principal questionnaire, student questionnaire, test results, portfolio analysis, peer questionnaire, and evaluator judgments). Debriefing sessions with evaluators and candidates provided additional insight into the functioning of the observation form in the classroom settings.

Data Source

During 1984-85, the observation form was put into use for general education teachers and field tested for use with special education and vocational education teachers. A total of 3,132 general education teachers were each observed three times using the same basic observation form. Observations were conducted by 105 general education teachers who had been trained as evaluators. The evaluators were trained for approximately 6 weeks to conduct the Career Ladder evaluations, including the observation, and worked for the Department of Education throughout the year.

All data collected by the evaluators were returned to the State Testing and Evaluation Center for processing and scoring. Most data, including the observation data, were handled on scannable documents. Individual items ("bubbles"), relating to each of the 17 observation scores incorporated in the scans (Section 2 and Section 3), were aggregated first within then across scans to generate an observation score for each of the three visits. The additional 8 items from Section 3 were available only at the observation visit level since they were coded only once for each visit.

Data from the three visits were then combined to generate the 25 overall observation scores. Each score was associated with one of the four domains measured on the observation: Planning (Domain I), Teaching Strategies (Domain II), Evaluation (Domain III), and Classroom Management (Domain IV). Three levels of analysis were, therefore, possible: by scan (18-27 scans per candidate), by visit (6-9 scans per visit, 3 visits per candidate), or per candidate.

Computation of the 25 scores varied according to the score. A sample of the observation form is provided in the Appendix. Each score was given a number and a short mnemonic name. These, and longer, more descriptive names are as follows:

Score	#	Description
TPREP	5	The teacher is prepared for class.
PACES	9	The teacher paces activity to accommodate learner understanding.
TASK	11	The teacher describes the task and content.
PROCQ	12	The students ask procedural questions. (Reverse Scoring)
LACKU	13	The students express/exhibit lack of understanding. (Reverse Scoring)
CINFO	14	The teacher provides correct information.
MONITOR	20	The teacher monitors student learning.
RETEACH	21	The teacher responds to lack of student understanding by reteaching.
REVWS	27	The teacher reviews.
SINIT	33	The students initiate or respond to interaction with the teacher.
ACADQ	34	The students ask academic questions.
ONTASK	35	The students are on task.
NONVOL	36	The teacher calls on non-volunteers.
SPREP	37	The students are prepared for class.
FEEDBK	46	The teacher provides students with feedback.
SANCTN	53	The teacher provides reasonable sanctions for inappropriate behavior.

CAUSE	54	Students misbehave prior to teacher action. (Reverse Scoring)
RESULT	55	Students misbehave after teacher action. (Reverse Scoring)
POSAFCT	60	The teacher displays positive affect toward students.
FAVOR	61	The teacher displays favoritism on the basis of race or sex. (Reverse Scoring)
MEDIA	67	The teacher uses media to enhance instruction.
TTTASK	68	Time is required to bring students back on task after an interruption. (Reverse Scoring)
RESRCE	69	The teacher makes use of available resources.
FACIL	70	The teacher makes use of available facilities.
IMPORT`	71	The teacher attends to needs in order of importance.

Data Characteristics

Descriptive statistics for the twenty-five variables based on the observation form are presented in Table 1. Most scores were skewed. Means for many variables tended to be near the maximum possible, including those for lack of bias (FAVOR), degree of student understanding (LACKU), percent of students ontask (ONTASK), and indices of off-task behavior (CAUSE, RESULT). These extreme averages were due to the way in which the variables were scores for this analysis and to the criterion-referenced basis on which the form was developed. It was important that teachers provide correct information and that they not show favoritism on the basis of race or sex. Therefore, these things were included even though they provided little measurement differentiation in identifying outstanding teachers.

Scores from some other variables tended toward the minimum possible. These included descriptions of task/content (TASK), monitoring (MONITOR), reviewing (REVWS), feedback (FEEDBK), student initiates or responds (SINIT), and positive effect (POSAFCT). Another factor which influenced scores was the inclusion of a set of teacher behaviors which together were expected to describe the activity of teaching. Teachers who scored high in one area were likely to have scored somewhat lower in the others, and it is unlikely that any teacher would have been performing the same activity (e.g., reviewing) on every scan on every observation.

Table 1

Descriptive Data for 25 Observations
Variables from General Education

1984-85 Evaluation Visits

SCORE #	SCORE NAME	DOMAIN	OVERALL MEAN	STANDARD DEVIATION	N of VISITS	MINIMUM SCORE	MAXIMUM SCORE
5	TPREP	I	98.6	11.9	8539	0	100
9	PACES	I	95.1	21.6	8031	0	100
11	TASK	II	17.8	8.5	8567	0	61
12	PROCQ	II	88.6	20.2	8567	0	100
13	LACKU	II	85.6	16.4	8567	0	100
14	CINFO	II	98.7	11.2	8515	0	100
20	MONITOR	II	22.7	9.7	8567	0	70
21	RETEACH	II	88.9	29.1	4461	0	100
27	REVWS	II	17.1	8.1	8567	0	64
33	SINIT	II	9.6	4.4	8567	0	30
34	ACADQ	II	54.8	37.0	8567	0	100
35	ONTASK	II	99.7	.6	8567	93	100
36	NONVOL	II	34.6	27.0	8567	0	100
37	SPREP	II	96.4	18.5	8001	0	100
46	FEEDBK	III	19.5	10.3	8567	0	59
53	SANCTN	IV	85.4	35.3	4192	0	100
54	CAUSE	IV	99.5	1.0	8567	88	100
55	RESULT	IV	99.8	.6	8567	93	100
60	POSAFCT	IV	2.4	2.1	8567	0	24
61	FAVOR	IV	100.0	.2	8567	91	100
67	MEDIA	IV	68.0	28.5	8567	0	100
68	TTTASK	IV	98.5	9.7	411	11	100
69	RESRCE	IV	94.6	17.9	8077	0	100
70	FACIL	IV	97.7	11.7	8439	0	100
71	IMPORT	IV	95.2	21.4	8079	0	100

Comparison to Field-Test Data

General education field-test data, collected in the spring of 1984, were based on a random sample of 275 teachers statewide. Actual evaluations were conducted during academic year 1984-85 on 3,132 teachers with a minimum of eight years of teaching experience who opted to join the program. It was expected that the combined effect of differences between the field and the actual evaluations would be a substantial increase in scores. Differences considered important were the use of volunteers, the minimum experience requirement, the additional preparation time, the additional information made available to candidates, and the incentive to do well in an actual evaluation.

Based on analysis of a sample of 2,947 candidates, substantially increased scores were generally the case. Eighteen of the scores increased, four decreased, two showed no large effects, and one could not be compared due to changes in scoring procedures.

Effect sizes were calculated as the difference between actual and field-test means divided by the actual standard deviation. The largest effect was related to an increase in the score for students on task with an effect size of 4.46. Reteaching also increased by more than a standard deviation. Effect sizes are presented in Table 2.

Decreases were associated with the amount of student involvement (student initiates/responds; calls on nonvolunteers), with positive affect, and with use of media. It is hypothesized that the decrease in student involvement and positive affect may have been due to the actual evaluation situation or to a perception that, since it was the teacher being evaluated, the maximum time possible should be devoted to teacher centered rather than student activities. The latter consideration may well have resulted in the decreased media use as well.

Table 2

Effect Sizes for Comparisons of Field Test With Actual Evaluation Scores - General Education Data

Score	Name	Effect
5	TPREP	.58
9	PACES	.20
11	TASK	.75
13	LACKU	.78
14	CINFO	.44
20	MONITOR	.74
21	RETEACH	1.70
27	REVWS	.03
33	SINIT	-.94
34	ACADQ	.15
35	ONTASK	4.46
36	NONVOL	-.83
37	SPREP	.40
46	FEEDBK	.69
53	SANCTN	.39
54	CAUSE	.28
55	RESULT	.06
60	POSAFT	-.45
61	FAVOR	.30
67	MEDIA	-.25
68	TTTASK	.14
69	RESRCE	.28
70	FACIL	.23
71	IMPORT	.19

Correlations Among Observation Variables

As anticipated, most correlations among observation visit variables in this analysis were low. This was expected for a number of reasons. First, most of the observation variables were severely skewed which could be expected to decrease correlations with other variables. Second, there was no a priori reason to expect high correlations between many of the variables. For example, large correlations were not expected between positive affect and presentation of correct information.

Third, the teacher behavior section was designed to be both responsive to a variety of teaching styles. Since it was important to allow for a variety of teaching styles (e.g., direct, indirect), different patterns of correlations might be expected to occur for different styles. In an overall analysis, these might not be detected.

In addition, a four minute scan was not long enough to permit the teacher to perform all of the behaviors which could be coded. Typically, a teacher would be seen to progress from one behavior (for example, presentation of content) to others (for example, reviewing) over a class period. This progression would be repeated on the other visits. This pattern would not be apparent in an overall analysis. Instead, one would expect to see low or even negative correlations among the teacher behaviors, since devoting a large amount of time to one would reduce the time available for another (additive model).

The highest correlation (.79) was between the CAUSE and RESULT variables. This was expected since both depended on the same incidents. These variables also had the largest number of positive correlations with other variables:

Variables	Correlation With	
	Cause	Result
SANCTN	.46	.47
IMPORT	.34	.35
TTTASK	.38	.34
ONTASK	.69	.72

A low level of student misbehavior in the classroom (CAUSE, RESULT, ONTASK) was related to rapid return to on-task behavior after an interruption (TTTASK), teacher attendance to needs in order of importance (IMPORT), and the use of appropriate sanctions for misbehavior (SANCTN). A causal direction is not explicit in the data. Teachers who attended to needs in order of importance and utilized appropriate sanctions for inappropriate behavior may have found, that this strategy decreased student off-task behaviors. Alternatively, teachers whose classrooms had relatively little off-task behavior may have had an easier time incorporating appropriate sanctions in their teaching and attending to needs in order of importance.

A second cluster of variables included positive affect, feedback, reviewing, monitoring, and student initiates/responds. In a third cluster, IMPORT (Teacher attends to tasks in order of importance) correlated with use of sanctions, student preparation, teacher preparation, pacing, and on-task behavior.

TTTASK had a number of nontrivial negative correlations. These were not interpretable because of the low numbers of observations for that variable. There was also a negative relation between task/content description (TASK) and reviewing (-.12). The teacher had a finite amount of time. More time spent in TASK resulted in less elsewhere. There were also negative relations between the students' level of understanding (LACKU) and FEEDBK (-.22) and MONITOR (-.18). When most children understood, there was less feedback and monitoring. Correlation coefficients are presented in Table 3.

Table 3

Correlation of Observation Scores
By Visit (N>8,500)

SCORE #	5	9	11	12	13	14	20	21	27	33	34	35	36	37	46	53	54	55	60	61	67	68	69	70	71
5	1.00																								
9	.29	1.00																							
11	.05	.06	1.00																						
12	.04	.05	.07	1.00																					
13	.00	.10	-.01	.07	1.00																				
14	.34	.25	.01	.00	-.02	1.00																			
20	.03	.09	.01	.00	-.18	.02	1.00																		
21	.08	.24	.12	.10	.00	.03	.13	1.00																	
27	.02	.05	-.12	-.06	-.07	.01	.38	.03	1.00																
33	.03	.05	.22	-.07	-.13	.02	.48	.11	.27	1.00															
34	.02	.03	.10	.41	-.13	.02	.07	.10	-.13	.02	1.00														
35	.13	.22	.14	.20	.10	.03	.15	.13	.07	.07	.06	1.00													
36	.03	.05	.12	.07	-.09	.02	.19	.05	.04	.15	.09	.10	1.00												
37	.36	.29	.04	.08	.06	.25	.09	.10	.05	.03	.02	.28	.03	1.00											
46	.03	.08	.11	-.03	-.22	.02	.42	.10	.25	.37	.08	.11	.26	.07	1.00										
53	.21	.32	.11	.08	.03	.20	.17	.10	.10	.13	.01	.44	.05	.30	.11	1.00									
54	.11	.20	.11	.28	.09	.03	.12	.13	.05	.03	.07	.69	.09	.23	.04	.46	1.00								
55	.11	.23	.11	.17	.08	.05	.12	.13	.06	.04	.05	.72	.09	.26	.05	.47	.79	1.00							
60	.04	.09	.16	.00	-.01	.01	.31	.08	.20	.27	.01	.15	.18	.08	.34	.14	.12	.12	1.00						
61	.01	.02	.00	-.01	.01	-.00	.01	.03	.02	.02	-.02	.02	-.01	.01	.02	-.01	.06	.05	.02	1.00					
67	.04	.03	.06	.02	-.05	.01	.06	.06	.04	.05	.04	.10	.06	.05	.03	.08	.06	.08	.08	.03	1.00				
68	-.01	-.02	-.12	-.05	.01	-.02	-.01	-.05	.05	-.06	-.10	-.03	.05	-.02	-.01	-.04	.38	.34	-.05	.58	.01	1.00			
69	.14	.13	.12	-.02	.04	.06	.04	.06	.05	.03	-.03	.12	-.01	.11	.05	.17	.10	.14	.11	.01	.17	-.05	1.00		
70	.09	.06	.00	.00	.01	.06	.01	.00	.02	.00	-.03	.05	-.01	.06	.04	.09	.05	.07	.03	.00	.04	.01	.08	1.00	
71	.33	.38	.08	.05	.04	.25	.09	.19	.03	.05	.05	.33	.04	.35	.04	.47	.34	.35	.08	-.01	.06	-.03	.08	.07	1.00

Results of Exploratory Factor Analysis

The factor structure of the observation data was explored using the more than 8,500 visits for which data were available. Correlation coefficients were calculated using pairwise deletion. Eight factors had eigenvalues greater than 1 and together accounted for 58% of the variance. The four-factor solution was selected based on the shape of the skree and is presented in Table 4.

The first factor is also the most stable, appearing in essentially the same form in solutions using from three to eight factors. It appears to be an index of management of student behavior in the classroom. Variables with high loadings included ONTASK, CAUSE, and RESULT.

Factor two relates to instructional organization, focusing on teacher and student preparedness, teacher providing correct information, pacing and IMPORT (Teacher attends to needs in order of importance) variables. The third factor relates to student-teacher interactions, with high loadings on teacher monitoring, reviewing, and feedback, student initiates/responds, pacing of instruction to accommodate learner understanding, and positive affect.

The fourth factor is difficult to interpret. It has high loadings for academic and procedural questions. It also has large negative loadings for favoritism (which had almost no variation due to extremely high scores), and for TTTASK for which much of the data was missing.

The factor structure provides partial support for the conceptual structure used in designing the observation form. The classroom management factor appears to be similar to Domain IV, Indicator A "Establishes and maintains appropriate learner behavior." The second and third factors combine items from several domains.

The second factor combines Domain I, Indicator B, "Prepares instructional plans and materials incorporating principals of effective instruction," with Domain II, Indicator A, "Provides a clear description of the learning task and its content," and Domain IV, Indicator A, "Establishes and maintains appropriate learner behavior." Each of these indicators relates to planning for and control of instruction. It is interesting to note that teacher attention to learner needs in order of importance appears to relate more closely to teaching strategies than to classroom management.

Factor three combines Domain II, Indicator B, "Monitors learner understanding and reteaches as necessary;" and Indicator D, "Establishes and maintains learner involvement in the learning task;" Domain III, Indicator A, "Uses information about learner performance to improve the instructional process;" and Domain IV, Indicator B, "Establishes and maintains a classroom climate conducive to learning." Each of these indicators relates to student/teacher interactions.

Interpretation of these factors must, of course, be tempered in light of the relatively small proportions of the variance explained by them.

Table 4

Factor Loadings Based on 4-Factor Solution of Observation Data (Data Based on Individual Visits)

VARIABLES	FACTOR 1	2	3	4
5 TREP	-.01	.66	.02	.01
9 PACES	.15	.62	.05	.07
11 TASK	.14	.04	.18	.37
12 PROCQ	.39	-.11	-.10	.52
13 LACKU	.14	.11	-.34	-.05
14 CINFO	-.11	.59	-.04	-.02
20 MONITOR	.08	.08	.75	-.01
21 RETEACH	.14	.18	.16	.22
27 REVWS	-.01	.11	.52	-.28
33 SINIT	-.03	.06	.71	.03
34 ACADQ	.18	-.15	.05	.59
35 ONTASK	.73	.25	.11	.18
36 NONVOL	.14	-.06	.38	.14
37 SPREP	.18	.61	.02	.05
46 FEEDBK	.02	.03	.72	.05
53 SANCTN	.44	.53	.14	.09
54 CAUSE	.89	.17	.03	-.01
55 RESULT	.86	.22	.05	-.03
60 POSAFCT	.10	.09	.56	.06
61 FAVOR	.28	-.11	.03	-.56
67 MEDIA	.10	.07	.13	.05
68 TTTASK	.48	-.17	-.04	-.69
69 RESRCE	.07	.31	.08	.00
70 FACIL	.02	.21	.01	-.07
71 IMPORT	.03	.63	.01	.13
Percent of Variance	16.2	9.6	7.6	6.6

Relationship of Scores from Different Visits

Correlations of observation variables from one visit to another were expected to be low for a wide variety of reasons. The three evaluators each observed the candidate once. Evaluator visits were spaced no less than 2 weeks apart, and usually occurred in the fall, winter, and spring (several months apart). Visits were conducted at varying times of day, including different content, and sometimes using different students (e.g., different high school classes). The first two visits were announced; the third, unannounced. After each visit, the evaluator provided the candidate with information on strengths and areas for improvement.

The correlations would, therefore, be attenuated not only by the skewed data previously discussed but also by any differences related to changes from one visit to another in occasion, content, student, and evaluator variables. Such changes may or may not be related to the reliability of the observation instrument. To the extent that observation differences are associated with actual differences in, for example, teaching style (for example, more emphasis on reviewing on one visit than another) or in classrooms (a less orderly class on one visit than another), they are entirely appropriate and do not indicate a deficiency in the instrument.

Although lack of correlation is not clearly interpretable, significant correlations across visits indicate stable patterns of teacher/classroom performance. Five variables showed such significant correlations, indicating performance was maintained across visits despite changes in evaluator, occasion, whether the visit was announced, students, subjects, and the effects of information provided on the previous visit. These exceptionally stable behaviors were: the amount of reviewing (REVWS), overall on-task behavior (ONTASK), incidents of off-task behavior (CAUSE), positive affect (POSAFCT), and use of media (MEDIA). Note that the relatively high correlations for time to task after interruptions (TTTASK) were nonsignificant due to the small number of observations for that variable. Correlations are presented in Table 5.

Table 5
Relationship of Scores From Different Visits

Score #	Name	Correlation Scores from Between Visits 1 & 2	2 & 3	1 & 3
5	TPREP	.03	.01	-.02
9	PACES	.04	.02	.07
11	TASK	.07	.10	.07
12	PROCQ	.18	.18	.19
13	LACKU	.05	.07	.07
14	CINFO	-.01	.02	-.01
20	MONITOR	.13	.16	.15
21	RETEACH	.02	.02	.11
27	REVWS	.20*	.23**	.21**
33	SINIT	.10	.10	.07
34	ACADQ	.18	.16	.17
35	ONTASK	.24**	.19*	.23**
36	NONVOL	.16	.15	.14
37	SPREP	.00	.05	.04
46	FEEDBK	.03	.08	.05
53	SANCTN	.13	.19	.14
54	CAUSE	.20*	.21**	.25**
55	RESULT	.13	.12	.17
60	POSAFCT	.24**	.22**	.19*
61	FAVOR	.00	.00	.00
67	MEDIA	.21**	.19*	.16
68	TTTASK	-.30	-.12	-.29
69	RESRCE	.07	.06	.08
70	FACIL	.10	.05	.13
71	IMPORT	.05	.09	.08

* $p < .05$
** $p < .01$

Conclusion

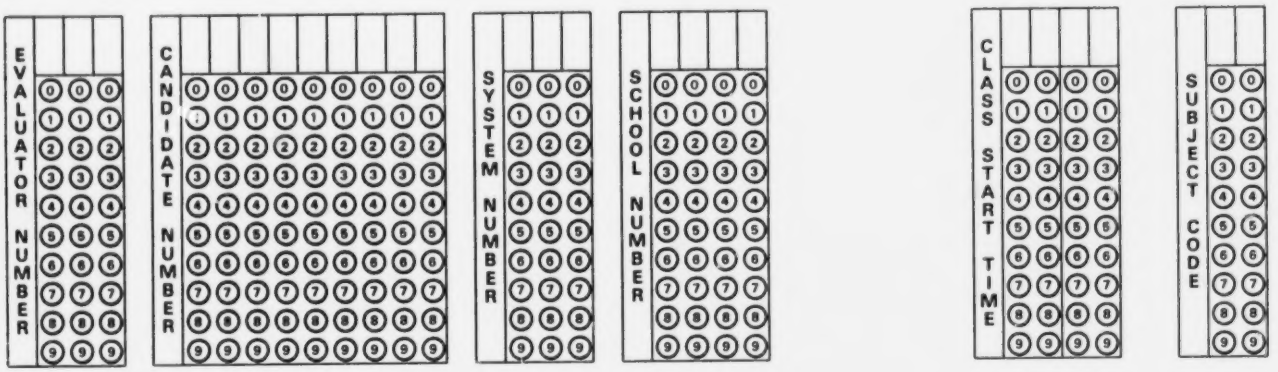
Data from this observation form are very complex. Scoring and analyses need to be undertaken and interpreted in light of this fact. Scans are nested in visits which are, in turm, nested in candidates. The choice of aggregation level determines, to some extent, the nature of the analytic outcome.

The choice of scoring method also exerts a powerful influence. The method reported here resulted in skewed distributions for most variables. It also made little use of the complex patterns of relationships among the data points within a scan and from one scan to the next within a visit. The tremendous potential of pattern scoring for these data is matched by the tremendous difficulty of developing meaningful scoring formulae of this type.

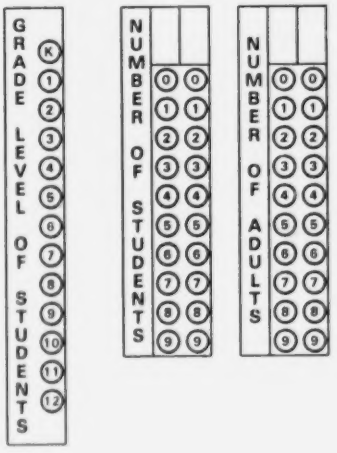
Additional work on understanding how this instrument functions in various contexts is also needed. The present analyses focused the interrelationships of the observation variables and on changes from the field test to actual evaluations and across three visits. These are only two of many contexts which should be considered in future analysis of these data.

TENNESSEE CAREER LADDER CLASSROOM OBSERVATION

ED-2510



CANDIDATE NAME _____



MARK IF ANY OF THE FOLLOWING CONDITIONS EXIST.

- ○ LOW SOCIO-ECONOMIC STATUS
- ○ ONE OR MORE MAINSTREAMED LEARNERS
- ○ SPECIAL GROUP; SLOW LEARNER
- ○ SPECIAL GROUP; GIFTED
- ○ LACK OF INSTRUCTIONAL MATERIALS
- ○ ASSIGNED TO INAPPROPRIATE FACILITIES
- ○ ASSIGNED INAPPROPRIATE COURSE/GRADE
- ○ ONE OR MORE LEARNERS WITH SPECIAL PROBLEMS; e.g., PHYSICAL DISORDER, EMOTIONAL PROBLEM
- ○ ORGANIZATIONAL DYSFUNCTION
- ○ UNANNOUNCED OBSERVATION
- ○ RELIABILITY CHECK

©1984, Career Ladder Program, Tennessee State Department of Education

NCS Trans-Optic® EB18-20002:321 A1407

OBSERVATION BEGAN AT: _____

OBSERVATION SCHEDULED TO END AT: _____

OBSERVATION ENDED AT: _____

		Teacher Behavior	Positive Affect	Student Initiates/Responds
TASK	Overview	○	○	○
	Directions	○	○	○
	Examples	○	○	○
	Distinguishes Steps	○	○	○
CONTENT	Definition/Examples	○	○	○
	Explanations/Rules/Attributes	○	○	○
	Provides Context	○	○	○
	Uses Learner Remarks	○	○	○
REVIEW	Initial Practice	○	○	○
	Independent Practice	○	○	○
	Varied Activities/Materials	○	○	○
	Group Practice	○	○	○
	Review	○	○	○
MONITORS	Requests Affirmation	○	○	○
	Requests Infor./Explanation	○	○	○
	Requests Comparison/Contrast	○	○	○
	Requests Evaluation	○	○	○
	Requests Questions	○	○	○
	Monitors Individual/Group	○	○	○
FEEDBACK	Recognizes Response	○	○	○
	Encourages Response	○	○	○
	Ignores Response	○	○	○
	Recognizes Correct Response	○	○	○
	Corrective Statement	○	○	○
	Corrected Assignments/Tests	○	○	○
	Reports Student Status	○	○	○
	Provides Recommendation	○	○	○
OTHER	Classroom Procedure	○	○	○
	Presentation/Other	○	○	○
	Teacher Off Task	○	○	○
	Greeting/Socialization	○	○	○

TEACHER CALLS ON STUDENTS WHO DO NOT VOLUNTEER ○

TIME ELAPSED BETWEEN THE TIME THE INTERRUPTION ENDED AND THE CLASS WAS BACK ON TASK IN MINUTES(M) & SECONDS(S)

M S S M S S M S S M S S M S S M S S

TEACHER CONTROL

CAUSE						TEACHER ACTION	RESULT					
STUDENT BEHAVIOR							STUDENT BEHAVIOR					
OFF TASK			DISRUPTIVE					OFF TASK		DISRUPTIVE		
INDIVIDUAL	SMALL GROUP	LARGE GROUP	INDIVIDUAL	SMALL GROUP	LARGE GROUP		COMPLIANCE	INDIVIDUAL	SMALL GROUP	LARGE GROUP	INDIVIDUAL	SMALL GROUP	LARGE GROUP
○	○	○	○	○	○	○	○	○	○	○	○	○	○

ACADEMIC PROCEDURAL

ACADEMIC _____

PROCEDURAL _____

PAGE 2

ADDITIONAL QUESTIONS

1. How many students are:

IN CLASS		OFF TASK	
0	0	0	0
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9

2. During most of the observation period, the teaching method was

- Lecture ○
- Discussion/Interaction ○
- Independent Work ○

3. During most of the observation period, the class was working:

- As a total group ○
- As subgroups ○
- As individuals ○

4. During the observation period, did students express/exhibit lack of understanding?

- None ○
- One ○
- Few (2-5) ○
- Many (more than 5) ○

5. Did the teacher respond to this by reteaching?

- Yes ○
- No ○
- N/A ○

6. Did you see evidence that the teacher showed favoritism to students based on

	Sex	Race
Academic Interaction	○	○
Discipline	○	○
Positive Affect	○	○

7. During the observation period, did the teacher use media or other classroom materials?

- Yes ○
- No ○

COMMENTS

POST–OBSERVATION QUESTIONS

		YES	NO	NOT OBSERVABLE
TEACHER COMMUNICATION	SPEAKS CLEARLY	○	○	○
	SPEAKS LOUDLY ENOUGH	○	○	○
	PACES SPEECH APPROPRIATELY	○	○	○
	ORGANIZES ORAL INFORMATION	○	○	○
	WRITES LEGIBLY	○	○	○
	USES CORRECT GRAMMAR	○	○	○
PLANNING/ MANAGEMENT	TEACHER PREPARED	○	○	○
	STUDENTS PREPARED	○	○	○
	TEACHER ATTENDED TO NEEDS IN ORDER OF IMPORTANCE	○	○	○
	TEACHER PACES ACTIVITY TO ACCOMMODATE LEARNER UNDERSTANDING	○	○	○
	TEACHER PRESENTS CORRECT INFORMATION	○	○	○
	TEACHER APPLIES REASONABLE SANCTIONS FOR INAPPROPRIATE BEHAVIOR	○	○	○
USE OF RESOURCES	MATERIALS/MEDIA ACCESSIBLE TO STUDENTS	○	○	○
	INTEGRATES RESOURCES WITH TEACHING	○	○	○
	PROVIDES SUPPLEMENTARY MATERIAL(S)	○	○	○
FACILITIES	CLASSROOM ACCOMMODATES SMALL GROUPS	○	○	○
	CLASSROOM ACCOMMODATES INDEPENDENT ACTIVITY	○	○	○
	FURNITURE PLACEMENT FACILITATES MOVEMENT	○	○	○